What factors determine whether a worker's trust in an AI system is well calibrated for a given task?

Shakhtar Donetsk
Politecnico di Torino
Type 3

Abstract

The influence of Artificial Intelligence (AI) on contemporary life is increasingly significant, reshaping how individuals approach various aspects of their professional activities. From university students to white-collar workers, AI tools are now widely used. Despite this widespread adoption, many aspects of AI remain unclear to users, particularly due to the relative novelty of the technology. As a result, many have not yet fully developed a stable and informed relationship with these systems. Since effective use of AI depends on appropriate levels of trust, understanding how such trust is formed and calibrated is essential for leveraging its full potential.

Keywords

Artificial Intelligence, AI, Trust in Technologies, Trust in Artificial Intelligence , Trust Calibration, Large Language Models, AI in Education

1 Introduction

Information technologies have become an inseparable part of everyday life, with recent data indicating that 97% of young people in the European Union use the internet daily [2,3]. More recently, Artificial Intelligence (AI), particularly Large Language Models (LLMs), has demonstrated the potential to significantly transform both daily and professional activities. Current estimates suggest that over half of U.S. adults have engaged with such systems [4]. As AI becomes increasingly integrated into workplace environments, it is influencing a wide range of tasks and decision-making processes. A key factor shaping the effective integration of AI is trust. Trust in AI systems determines whether users choose to rely on, ignore, or critically evaluate algorithmic recommendations [5, 6]. However, trust is not inherently beneficial: both overtrust and undertrust can negatively impact performance [5, 6]. Overtrust, often described as automation bias, leads users to accept AI outputs without sufficient scrutiny [7]. Undertrust, or algorithm aversion, leads users to reject useful AI assistance even when it outperforms human judgment [8]. This highlights the importance of trust calibration, defined as the alignment between a system's actual capabilities and the user's actual level of trust [5].Factors such as transparency and explainability have been shown to significantly influence user trust in AI systems [9, 10]. Cultural background, prior experience, and perceived risk further shape how users interact with AI [11, 12]. Despite growing research in this area, important questions remain regarding how trust is established, how it varies across contexts, and under what conditions it is appropriately calibrated. This study aims to map the existing literature on trust in AI, with a particular focus on identifying the factors that influence trust formation and examining the conditions under which trust is well calibrated for a given task.

2 Method

To address the research objectives, a twofold methodological approach was adopted, combining a structured scoping review of the literature with an empirical survey.

2.1 Literature Review

A scoping review was conducted following the PRISMA-ScR guidelines found in section D of the Appendix. Three academic databases were searched: ACM Digital Library, Scopus, and Google Scholar. The primary search strings used were: "human trust in AI", "trust calibration artificial intelligence", "automation bias", "algorithm aversion", "AI transparency", and "AI in education". The publication window was restricted primarily to 2015–2026 to reflect recent developments in AI, though foundational works published before this period were retained where their theoretical contribution remained central to the field (notably Lee & See, 2004 [5] and Parasuraman et al., 2000 [7]). Each group member independently reviewed up to ten candidate articles and recorded them in a shared tracking table. An initial pool of 60 articles was identified. After duplicate removal, 44 unique sources remained. These were then screened by title and abstract against the following inclusion criteria: (i) empirical or theoretical focus on trust in automated or AI systems; (ii) peer-reviewed or reputable grey literature; (iii) written in English. Sources were excluded if they focused exclusively on human-to-human trust without reference to automation, or if they addressed AI purely from a technical performance standpoint without reference to user behavior or trust. This process yielded a final set of 19 sources used in the review.

2.2 Empirical Survey

An empirical survey was conducted to complement the literature findings with primary data. The survey collected responses from

over 100 participants, primarily university students, and was structured into three sections: (1) demographic information, (2) general attitudes toward AI and trust, and (3) perceptions of AI integration in university teaching. The questionnaire included Likert-scale and multiple-choice questions to capture both attitudes and behavioral tendencies. As the sample consists primarily of university students, results are not generalizable to the broader working population. Survey findings are therefore used to contextualize and illustrate the literature, not to test hypotheses. Further details on survey design and results are provided in the Appendix.

3 Framework

Trust in Artificial Intelligence is a multi-dimensional concept shaped by the interaction between system characteristics, human factors, and the broader context in which the technology is used [5, 6]. Rather than being determined by a single element, trust emerges from a combination of technical performance, user perceptions, and external conditions such as regulation and application domain [12, 13]. This framework organizes these influences into five categories. It is designed to be domain-agnostic; applications to specific domains are illustrated in Section 3.7.

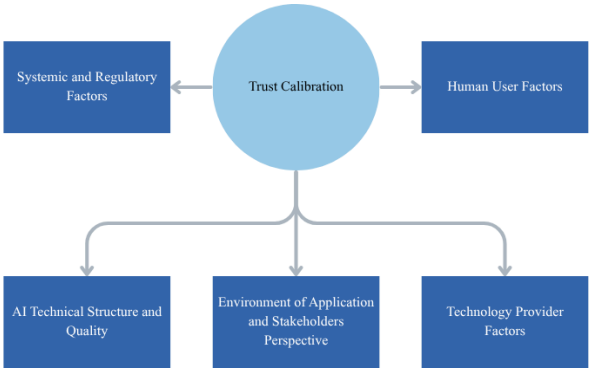

Fig. 1 Framework Diagram

3.1. AI Technical Structure and Quality

Trust in AI systems is strongly influenced by their technical performance and reliability. A key determinant is reliability over time, defined as the system's ability to produce consistent and accurate outputs across repeated interactions. Empirical research shows that users adjust their level of reliance based on perceived system performance, making reliability a central component of trust formation [6, 8]. Importantly, reliability is often task-dependent: users may trust AI systems in routine or well-defined tasks, while remaining more cautious in complex or ambiguous situations [6]. This indicates that trust is not uniform, but varies depending on how users evaluate the demands and risks of specific tasks. A second technical factor is explainability, which refers to the system's ability to provide understandable justifications for its outputs. Systems that generate explanations aligned with human reasoning processes are more likely to be perceived as trustworthy, even when their underlying mechanisms remain complex [9, 10]. Despite these strengths, technical limitations such as hidden biases or inconsistent performance can lead to miscalibrated trust, as users may not always accurately perceive system limitations [14, 15]. While Hoff and Bashir [6] emphasize system reliability as a primary driver of trust, Glikson and Woolley [11] argue that user perception may outweigh objective performance, suggesting that trust cannot be explained by technical factors alone.

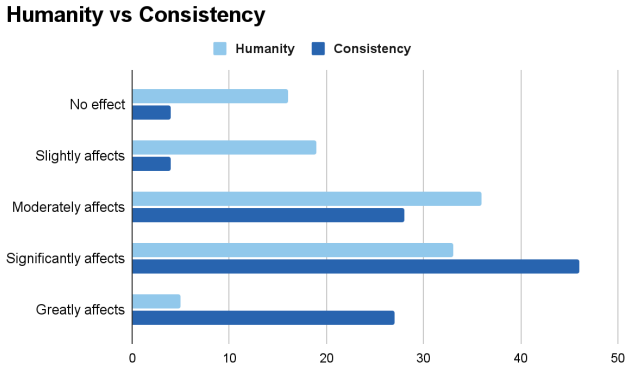

Fig. 2 Comparison of the influence of perceived reliability and alignment with human reasoning (humanity) on user trust (survey results)

3.2. Human User Factors

Trust is also shaped by characteristics of the user. Research shows that factors such as experience, education, and familiarity with AI systems significantly influence how users interpret and rely on AI outputs [6, 11]. More experienced users tend to develop more calibrated trust, adjusting their reliance based on context rather than treating AI systems as uniformly reliable or unreliable. Cultural background has also been proposed as a factor influencing baseline attitudes toward AI, potentially affecting expectations and willingness to trust automated systems [12], though this claim currently rests on limited empirical evidence, as the only source reviewed was a small workshop study; specifically, cross-cultural comparative research on trust calibration remains a significant gap in the literature. In addition, personal involvement with a task plays a critical role. When decisions have direct consequences for the user, perceived risk increases, often leading to more cautious behavior and reduced reliance on AI [13]. These findings suggest that trust is not solely determined by system performance, but also by how users

perceive risk and personal impact in a given situation. A notable behavioral pattern in this category is algorithm aversion, the tendency to distrust and avoid algorithmic outputs even when the algorithm demonstrably outperforms human judgment [8]. This is distinct from, but related to, undertrust: algorithm aversion can persist even after a user is shown evidence of the algorithm's superior accuracy, particularly when the user observes the algorithm make any error at all.

3.3. Systemic and Regulatory Factors

Trust in AI is also influenced by the broader institutional and regulatory environment. Laws and governance frameworks provide external assurances regarding system safety, fairness, and accountability, thereby shaping user confidence [16]. A central issue in this context is accountability: when responsibility for AI-assisted decisions is unclear, whether attributed to the user, the system, or the deploying organization, users may be less willing to rely on AI outputs [13]. This lack of clarity can undermine trust even when systems demonstrate high technical performance. To address this challenge, many systems incorporate mechanisms, where human oversight is maintained over critical decisions. Such approaches help balance automation with accountability, supporting more appropriate trust relationships [17]. Regulatory frameworks such as the EU AI Act, formally adopted by the European Parliament in March 2024 and progressively entering into force through 2025–2027, represent an important step toward formalized accountability, though their practical effect on user trust calibration in everyday professional settings remains an area requiring further empirical investigation [16].

3.4. Technology Provider Factors

Trust is also shaped by the practices of organizations that develop and deploy AI systems. Technology providers influence trust through internal decisions including system design, transparency practices, and ethical standards. Transparency regarding training data, system limitations, and decision processes plays a key role in shaping user perceptions [18]. A lack of transparency may lead users to suspect hidden biases or unclear intentions, reducing trust even in technically capable systems. Closely related is the issue of ethical alignment, which concerns the principles embedded within AI systems. Since ethical standards vary across organizations and contexts, determining which values are reflected in AI behavior remains complex [14, 15]. This uncertainty can further influence how users perceive and trust AI systems, particularly when users are aware that a system has been designed with commercial rather than user-centered priorities. Note that while this category and Category 3.3 both touch on accountability, they are analytically distinct: regulatory factors are external to the provider (imposed by law), while provider factors concern internal organizational decisions and values that exist independently of regulation.

3.5. Environment of Application and Stakeholders Perspectives

Trust varies significantly across application contexts, reflecting differences in risk, domain characteristics, and stakeholder involvement.

3.5.1 High vs. Low Stakes. In low-stakes environments, users are generally more willing to rely on AI systems and experiment with their capabilities. In contrast, high-stakes situations lead to increased scrutiny and more cautious reliance, even when system performance remains unchanged [13]. This asymmetry has important design implications: systems used in high-stakes environments may require additional transparency features or human oversight mechanisms to achieve appropriate trust levels.

3.5.2 Domain Type (Technical vs. Human-Centered Fields). Trust also differs across domains. Users tend to exhibit higher trust in technical or objective domains, where outcomes are perceived as measurable and predictable. Trust is often lower in human-centered domains such as healthcare, education, or hiring, where decisions require contextual judgment and ethical consideration [11].

3.5.3 Stakeholder Perspectives. Trust is further influenced by the perspectives of different stakeholders involved in the AI ecosystem. Each stakeholder group faces distinct risks and incentives, which shape how AI systems are developed, deployed, and governed.

Stakeholder	**Role in the AI Ecosystem**	**Primary Perceived Risk**
Worker	Primary decision maker	Overtrust & Automation Bias, Undertrust, Skill degradation
Organization/ Employers	Deploy AI system	Liability, Reduced trust, Operational risk.
Customer	Indirectly affected by AI	Bias & Unfair outcomes, Lack of transparency
Policy Maker	Establish safety guidelines for AI use	Insufficient or lagging regulation, Erosion of public trust.

Table 1: Stakeholders and Primary Perceived Risk

3.6. Cross-Cutting Issues: Trust Calibration and Risks

The five categories described above each contribute independently to trust formation, but their most consequential interaction is in determining whether trust ends up well matched to what a system can actually do for a specific task. This matching problem trust calibration is the central challenge the framework is designed to address. Miscalibration in either direction carries real costs: a worker who trusts too much forgoes the critical judgment that catches AI errors; a worker who trusts too little discards useful assistance and may perform worse than if no AI were available at all. The following subsections identify the three cross-cutting forces that most commonly produce miscalibration regardless of which category dominates in a given context.

3.6.1 Trust Miscalibration. Miscalibration arises when a user's level of trust diverges from the system's actual capability on a given task. The divergence can run in either direction. Overtrust commonly called automation bias leads users to accept outputs without applying independent judgment, even in conditions where errors are consequential and detectable [6]. The less-studied but equally damaging direction is undertrust, where users reject or systematically discount AI outputs despite the system outperforming human judgment on that task class; critically, this rejection can persist even after a user has been shown accuracy data, particularly if they have witnessed a single salient failure [8]. Both failure modes share a common structural cause: users evaluate AI systems holistically rather than task-specifically, generalizing from performance in one domain to unwarranted conclusions about another. This is precisely why miscalibration is treated here as a cross-cutting issue rather than a property of any single framework dimension it can be produced by technical opacity [3.1], user inexperience [3.2], unclear accountability [3.3], or high-stakes context [3.5.1] acting individually or in combination.

3.6.2 Organizational and Economic Pressures. In real-world settings, trust is not formed solely through individual judgment. Organizational incentives such as efficiency goals or competitive pressures may encourage increased reliance on AI systems even when such reliance is not fully justified [19]. This can lead to reduced critical evaluation and increased risk of overreliance. Raisch and Krakowski [19] describe this as the automation–augmentation paradox: organizations systematically favor automation over human augmentation, creating structural conditions for miscalibrated trust.

3.6.3 Bias and Ethical Risks. AI systems may reflect or amplify biases present in their training data, leading to unfair or discriminatory outcomes [14]. These risks emerge across multiple levels; system design, organizational practices, and deployment contexts and represent a cross-cutting issue that directly impacts trust, particularly in sensitive or high-stakes applications. Users who encounter biased outputs, or who are aware that a system may be biased, will rationally adjust their trust downward; the challenge is that such adjustments are often not well-calibrated either, leading to either overcorrection (rejecting useful systems) or undercorrection (continuing to rely on biased outputs).

3.7. Synthesis: The Calibration Continuum

While this framework identifies distinct categories of influence, they converge into a "calibration continuum" that helps bridge the gap between the university-based survey data and the professional white-collar environments identified as a literature gap.

The Literacy Bridge: The survey suggests that even without formal training, users inherently adjust trust based on perceived stakes. In a professional context, this implies that "AI literacy" is not just technical knowledge, but the ability to recognize which task-class a specific AI output belongs to.

Recovery of Trust: A critical cross-cutting issue is how calibration evolves after a failure. Professionals, much like the "skeptical" students in high-stakes grading scenarios, may default to permanent algorithm aversion after a single error if accountability structures are not clearly defined by the provider or the organization.

Structural Proxies: Although white-collar workers are underrepresented in the current corpus, the "Professor Use Case" serves as a structural proxy; the professor’s shift from "passive acceptor" to "active evaluator" via explainability features illustrates how professional calibration must move beyond holistic "gut feelings" toward task-specific evidence.

The Role of Institutional Signals: The transition from low-stakes experimentation to high-stakes reliance is often triggered by institutional endorsement. If an organization mimics the university's endorsement of AI tools, it risks creating a "default trust" environment where social proof replaces conscious, task-specific evaluation.

4. Worked Use Case: University Professor

To put into perspective how the framework’s dimensions are employed in practice, consider a university professor who integrates AI tools into their teaching and assessment workflow. This use case is particularly relevant given the survey data collected in this study.

4.1 Task Mapping: Where Trust Demands Differ

A professor’s work includes tasks with different trust requirements.

4.1.1 Drafting lecture content. When using an AI tool to generate first drafts of slideshows or explanatory texts, the professor remains at the final checkpoint before the material reaches the students. The stakes are low as errors are detectable and the cost of the occasional AI failure is recoverable. The primary long-term potential downside of this method is the skill degradation of the professor as it will possibly erode the professor's own capacity to form complex ideas due to a high degree of reliance on the AI.

4.1.2 Providing feedback on student assignment AI-generated feedback could be fluent and structurally well done while missing disciplinary nuance and possibly misreading a student's argument. The stakes of the usage are medium as this task is a human-centered domain and uncritical reliance carries a real risk. Therefore, automation bias under time pressure and accountability is also implicated as students hold the expectation that their work was evaluated by a qualified professor.

4.1.3 Detecting academic dishonesty This is a high-stakes decision with direct consequences for an individual student. AI detection tools are known to produce both false negatives and false positives in their output and can reflect biases related to writing style or disciplinary convention. In this scenario, deliberate skepticism toward AI outputs is not miscalibration and it is the appropriate response as the errors fall on a third party.

4.2 Key Interactions

4.2.1 Reliability is task-specific, not global. A professor who finds an LLM reliable for drafting should not infer the same reliability for academic dishonesty detection. These are structurally different tasks, and strong performance on one doesn't enforce performance about the other.

4.2.2 Explainability shifts the nature of reliance. When the AI feedback tool underlines which rubric criteria it applied, the professor is repositioned from passive acceptor to an active evaluator role. This is precisely, where the explainability offers a practical value: not as the trust building feature, but rather as a mechanism that makes critical engagement possible with every output.

4.2.3 Cultural and institutional signals shape baseline attitudes If a university publicly endorses specific AI tools, or if colleagues visibly rely on them without apparent consequence, individual professors receive a signal that high reliance is tolerated. These social and institutional constraints can shift baseline trust independently of any direct experience with the system and it will allow a mechanism to exist without conscious evaluation allowing reliance to develop without conscious evaluation of the system's actual track record. This mirrors the institutional dimension identified in section 3.3: when accountability structures are absent or diffuse, trust defaults to social proof rather than task-specific evidence

4.3 What Well-Calibrated Trust Looks Like Here

A professor with well-calibrated trust systematically applies different levels of reliance across tasks: high trust with critical review for content drafting, active scrutiny for AI-assisted feedback, and deliberate skepticism for dishonesty detection. This differentiation is not accidental; it reflects the joint operation of three framework dimensions. Technical reliability [3.1] is treated as task-specific rather than global. Domain type [3.5.2] raises the scrutiny threshold in human-centered tasks where contextual judgment matters. And accountability awareness [3.3] drives deliberate rejection of AI authority where errors fall on a third party. Calibration of this kind is also not static: it should be revised as the professor accumulates direct evidence of where the system succeeds and fails, maintaining a task-sensitive orientation throughout.

4.4 Survey Findings

The survey findings, collected from more than 100 university students, provide primary data support for the claims developed in the professor use case.

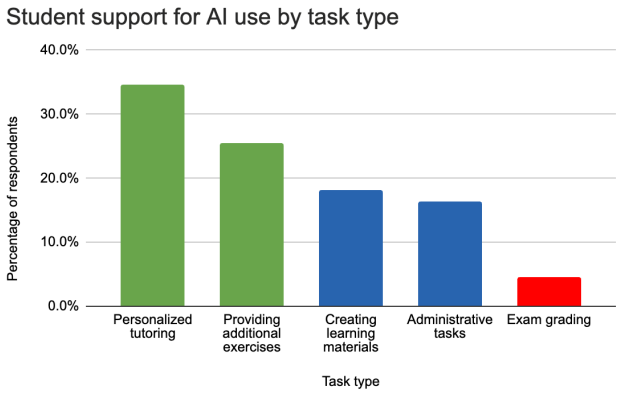

Fig. 3 Student support for AI use by task type ordered by perceived stakes (Lower, Medium, Higher)

Figure 3 presents the distribution of student preferences across five potential applications of AI in university teaching. The results create a noticeable pattern of task-sensitive trust where students strongly endorse AI use in personalized tutoring and exercise generation, tasks that have low-stakes. Endorsement falls progressively as the stakes of the task increase with exam grading selected by a negligible minority. Furthermore, students seem to grasp the distinction between low-stake tasks and high-stakes tasks which they treat with skepticism, suggesting that perceived stakes are salient even without formal training in trust calibration.

Figure 4 adds a further dimension to the whole picture. The majority of the respondents expressed optimism when they were asked whether AI can improve the overall quality of university teaching. This is a meaningful result because it suggests that student reluctance towards specific high-stake application does not reflect a general rejection of AI in educational context. This concept is where well-calibrated trust relies as AI can be introduced where its capabilities are relevant and its errors are recoverable, while not relying fully where accountability and fairness are at stake.

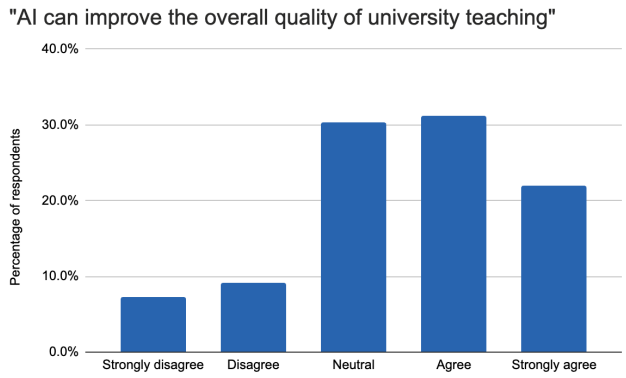

Fig. 4 Student attitudes toward AI's potential to improve university teaching quality

5. Limitations and Future Work

Several specific gaps limit the framework's generalizability.

Provider-side and regulatory factors lack behavioral evidence. Transparency and ethical alignment are theoretically argued [15, 16, 18] but no reviewed study directly measures how users' trust responds to disclosures about training data or model limitations. Controlled behavioral experiments are needed. Empirical foundations are concentrated in high-automation domains. The core reliance studies [5, 7, 13] were conducted in aviation, military, and process control. Their transferability to knowledge work and LLM-assisted professional tasks has not been tested. White-collar worker populations are almost entirely absent from the reviewed corpus. Cultural factors rest on thin evidence. Cultural background is identified as a relevant user factor [12], but the sole empirical source is a small workshop paper. Cross-cultural comparative studies examining whether and how national or organizational culture affects calibration specifically do not exist in the reviewed literature.

Longitudinal trust dynamics are unaddressed. Every empirical source examined trust at a single point or over very short periods. How calibration evolves with accumulated experience and whether it recovers after observed errors remains an open empirical question [5, 6]. Undertrust is underrepresented. Most empirical work focuses on automation bias. The conditions under which undertrust dominates, and whether specific design choices can address algorithm aversion [8], have not been isolated for AI systems specifically. Organizational pressure on individual calibration has no empirical grounding. The automation–augmentation paradox [19] is a management-level construct. How institutional incentives concretely shape individual reliance decisions in practice has not been studied behaviorally.

References

[1] Student Survey on Trust in Artificial Intelligence. 2026. Conducted as part of this study; see ZIP Archive for methodology and results.

[2] Eurostat. 2025. 97% of young people in the EU use the internet daily. https://ec.europa.eu/eurostat/web/products-eurostat-news/w/edn-20250715-1

[3] Eurostat. N.d. Individuals using the internet daily – Data browser. https://ec.europa.eu/eurostat/databrowser/view/tin00028/default/table?lang=en

[4] Elon University. 2025. Survey: 52% of U.S. adults now use AI large language models like ChatGPT. https://www.elon.edu/u/news/2025/03/12/survey-52-of-u-s-adults-now-use-ai-large-language-models-like-chatgpt

[5] John D. Lee and Katrina A. See. 2004. Trust in automation: Designing for appropriate reliance. Human Factors: The Journal of the Human Factors and Ergonomics Society 46, 1, 50–80. doi:10.1518/hfes.46.1.50_30392

[6] Kevin A. Hoff and Masooda N. Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. Human Factors: The Journal of the Human Factors and Ergonomics Society 57, 3, 407–434. doi:10.1177/0018720814547570

[7] Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens. 2000. A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics — Part A: Systems and Humans 30, 3, 286–297. doi:10.1109/3468.844354

[8] Jason W. Burton, Mari-Klara Stein, and Tina Blegind Jensen. 2020. A systematic review of algorithm aversion in augmented decision making. Journal of Behavioral Decision Making 33, 2, 220–239. doi:10.1002/bdm.2155

[9] Finale Doshi-Velez and Been Kim. 2017. Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

[10] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence 267, 1–38. doi:10.1016/j.artint.2018.07.007

[11] Ella Glikson and Anita Williams Woolley. 2020. Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals 14, 2, 627–660. doi:10.5465/annals.2018.0057

[12] Masooda Bashir and Hsiao-Ying Huang. 2017. Cultural influences on the trustworthiness of conversational agents. In Proceedings of the 2017 ACM Workshop on An Application-Oriented Approach to BCI Out of the Laboratory (Brighton, United Kingdom) (BCIforReal '17). ACM, New York, NY, USA, 23–27. doi:10.1145/3125739.3125749

[13] Mary T. Dzindolet, Scott A. Peterson, Regina A. Pomranky, Linda G. Pierce, and Hall P. Beck. 2003. The role of trust in automation reliance. International Journal of Human-Computer Studies 58, 6, 697–718. doi:10.1016/S1071-5819(03)00038-7

[14] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A survey on bias and fairness in machine learning. ACM Computing Surveys 54, 6, Article 115, 35 pages. doi:10.1145/3457607

[15] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. Nature Machine Intelligence 1, 389–399. doi:10.1038/s42256-019-0088-2

[16] Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, Peggy Valcke, and Effy Vayena. 2018. AI4People — An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines 28, 4, 689–707. doi:10.1007/s11023-018-9482-5

[17] European Parliament and Council of the European Union. 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. Official Journal of the European Union L 2024/1689. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_2024_1689

[18] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for human-AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland, UK) (CHI '19). ACM, New York, NY, USA, Article 3, 13 pages. doi:10.1145/3290605.3300233

[19] Sebastian Raisch and Sebastian Krakowski. 2021. Artificial intelligence and management: The automation–augmentation paradox. Academy of Management Review 46, 1, 192–210. doi:10.5465/amr.2018.0072

A ZIP Archive

The Google Drive link provides access to a ZIP archive containing the README.txt documentation, the Source-table.csv literature list, and the AI Trust calibration survey.csv. These files provide transparency regarding our paper selection process, detail our coding methodology, and ensure the reproducibility of our research: https://drive.google.com/file/d/1XMRXqmbA9Y7QiYfkM5H_lZYxNVMCs8Jh/view?usp=sharing

B Team Members of Shakhtar Donetsk

Table 2: Team Members' name, surname and institutional email

Authors		
Name	**Surname**	**Email**
Nikola	Kandev	s323849@studenti.polito.it
Petru	Nacea	s321901@studenti.polito.it
Nicolò	Lodigiani	s340758@studenti.polito.it
Nicolò	Moiso	s327721@studenti.polito.it
Riccardo	Ferrero	s336535@studenti.polito.it
Tommaso	Palena	s325536@studenti.polito.it
Davide	Pandino	s336928@studenti.polito.it
Pietro	Montorsi	s326382@studenti.polito.it

Leonardo	Vezzù	s343869@studenti.polito.it

C AI Use Disclosure

All AI tools used during this project were limited to assistance with phrasing and initial orientation during the literature search. Specifically, ChatGPT and Gemini were consulted to suggest alternative formulations for selected passages and to generate preliminary summaries of candidate sources during the early screening phase.

All sources cited in this paper were read in full by at least one team member and cross-checked by a second. No citation was included on the basis of an AI-generated summary alone. AI-suggested phrasings were reviewed and either rewritten or explicitly approved by the team members responsible for each section before inclusion. The final text of all sections reflects the team's own analysis and judgment.

D PRISMA-ScR flow diagram